



NAMRL-1404

**PROBABILITY OF SUCCESS IN PRIMARY
FLIGHT TRAINING AS A FUNCTION OF
ASTB SCORES AND API GRADES: AN
EXAMPLE OF THE STATISTICAL
INFERENCING COMPONENT OF THE
PILOT PREDICTION SYSTEM**

D. J. Blower

**Naval Aerospace Medical Research Laboratory
51 Hovey Road
Pensacola, Florida 32508-1046**

DTIC QUALITY INSPECTED 4

Approved for public release; distribution unlimited.

Reviewed and approved 30 Nov 98

C. G. Armstrong

C. G. ARMSTRONG, CAPT, MSC USN
Commanding Officer



This research was sponsored by the Naval Medical Research and Development Command under work unit 62233NMM33P30.001-7801.

The views expressed in this article are those of the authors and do not reflect the official policy or position of the Department of the Navy, Department of Defense, nor the U.S. Government.

Volunteer subjects were recruited, evaluated, and employed in accordance with the procedures specified in the Department of Defense Directive 3216.2 and Secretary of the Navy Instruction 3900.39 series. These instructions are based upon voluntary informed consent and meet or exceed the provisions of prevailing national and international guidelines

Trade names of materials and/or products of commercial or nongovernment organizations are cited as needed for precision. These citations do not constitute official endorsement or approval of the use of such commercial materials and/or products.


Reproduction in whole or in part is permitted for any purpose of the United States Government.

**NAVAL AEROSPACE MEDICAL RESEARCH LABORATORY
51 HOVEY ROAD, PENSACOLA, FL 32508-1046**

NAMRL-1404

**PROBABILITY OF SUCCESS IN PRIMARY FLIGHT TRAINING AS A
FUNCTION OF ASTB SCORES AND API GRADES: AN EXAMPLE OF
THE STATISTICAL INFERENCING COMPONENT OF THE PILOT
PREDICTION SYSTEM**

D. J. Blower


19990303 169

ABSTRACT

The Pilot Prediction System (PPS) is a research effort designed to provide Navy managers and other decision makers with improved access to selection and training data. Many disparate data bases, each containing partial and sometimes overlapping information on selection data and training performance, currently exist. There has been no attempt to coordinate the bits and pieces gathered into these local databases into a coherent whole. Such data needs to be merged and the anomalies excised so that a more global picture of selection data and training performance can emerge. In addition, the targeted users of the PPS should be shielded from the low-level technicalities of the data base because such technical details are of no concern to them. For the same reason, the statistical manipulations that provide extrapolations from the data base to new cases can be hidden from view. This report documents the first efforts at constructing the statistical modeling component of the PPS as derived from Bayesian statistical decision theory. It enables the potential user of the PPS to predict success in primary flight training for flight students based on their scores on selection tests and ground school performance.

Acknowledgments

I would like to acknowledge the support of LT Henry Williams, Head of the Aviation Selection Division at the Naval Aerospace Medical Research Laboratory (NAMRL), who provided the data for this report. Thanks also go to my fellow members of the PPS team at the time this report was written: LCDR Sean Biggerstaff, LT Henry Williams, Mr. Allen Chapman, and Mrs. Claire Portman-Tiller.

INTRODUCTION

The Pilot Prediction System (PPS) is a research effort designed to provide Navy managers and other decision makers with improved access to selection and training data. In the past, our collaborative work with the Chief, Naval Education and Training (CNET) has pinpointed certain deficiencies in providing technical support for high-level policy questions concerned with selection and training issues.

Many disparate data bases, each containing partial and sometimes overlapping information on selection data and training performance, currently exist. The data in these databases are often structured according to different software formatting conventions. The proprietors of these data bases are geographically dispersed and under the control of different administrators. The quality of the data varies greatly. There has been no attempt to coordinate the bits and pieces gathered into these local databases into a coherent whole. These disjunctive data need to be merged and the anomalies excised so that a more global picture of selection data and training performance can emerge. When managers need to make a decision based on training data they should have to consult only one comprehensive, error-free data base.

In addition, the targeted users of the PPS should be shielded from the low-level technicalities of accessing the data base as well as the statistical manipulations that provide extrapolations from the data base to new cases. One of the PPS's major goals is to provide access to an improved data base in a "user-friendly" manner. One analogy that we employ quite often in describing the PPS is that it should function much like a typical spreadsheet. This would allow users to conduct "what-if" analyses on scenarios of their choosing. This spreadsheet would operate at the user's desired level of what is important and keep, what is for them, the unimportant technical operations hidden from view.

A primary consideration of this project is to provide statistical models that allow decision makers to take full advantage of the available data in order to make predictions about the future success of an individual. A map or schematic of the overall selection and training flow would be shown to the user on system start-up. The potential user of the system need only indicate via a mouse click which available data connected to the map should go into a model as predictor variables. Likewise, a mouse click would indicate which criterion variable is desired. The appropriate section of the map would then be highlighted to show the user what portion of the overall schema he has chosen to investigate. The user could then indicate which candidate he is interested in looking at. The system would automatically fetch the required data for this individual from the data base.

The PPS would link these choices into a statistical model. The numerical routines in the software would then perform an inference indicating the predicted outcome for this individual on the chosen criterion and then, most importantly, also provide the degree of confidence that the user might have for this prediction by the system.

For example, scores on the various subtests of the Aviation Selection Test Battery (ASTB) and the overall Navy standard score from Aviation Preflight Indoctrination (API) could be the predictor variables of interest. Graduation or attrition from primary flight training could be the criterion of interest. Perhaps a student comes before a Progress Review Board because of substandard performance during scheduled early familiarization flights in primary flight training. Entering the student's Social Security Number into the system would immediately access his/her ASTB scores and overall performance during API from the data base and enter them into the statistical model. The PPS would then show whether this student was a predicted pass or a predicted fail from primary flight training. In addition, the confidence level of this prediction would be indicated by giving the probability of this prediction being true. This student might be a predicted pass given his/her pattern of scores, but the probability of the predicted pass would be only 60%. A threshold confidence level might have been established at, say, an 80% confidence level with the result that this student might be moved into the predicted fail category.

In this fashion we envision the PPS as being useful to the Commodore of the Training Wing where this student was undergoing training. He or she might use the PPS to help make an informed decision about the potential success of this student when compared to the success profiles of the students already in the data base.

The Observed Data

The data analyzed in this report were provided by LT Henry Williams. The ASTB test scores were compiled at the Naval Operational Medicine Institute (NOMI), and the API scores came courtesy of the Naval Schools Command (NASC), both at NAS Pensacola. The primary flight training attrition data were collected from Training Wing 5, which consists of three training squadrons, VT-2, VT-3, and VT-6, located at Milton, Florida. The primary flight training attrition data covered the time span from June 1994 to January 1998. The 1,054 entries in the data base were composed exclusively of Navy and Marine Corps student pilots. Only those students who had taken the version of the ASTB as revised in 1992 were included in the data base. A total of 94% of the students were male and the remaining 6% female. From the overall total of 1,054 students in the data base, 936 students passed primary flight training and 118 students were attrited (i.e., failed training).

The analysis in this report is predicated on the assumption of multivariate normality for the predictor variables. This section lists the sufficient statistics for this assumption, viz., the sample means and the sample variance-covariance matrix. Table 1 presents the means and standard deviations for the five predictor variables studied in this report. MVT, MCT, SAT, and ANI are four subtests from the ASTB. MVT stands for Mathematics

Table 1: Means, standard deviations, and sample sizes over the five predictor variables for the two categories of pass and fail in primary flight training.

Test	Pass			Fail		
	Mean	SD	N	Mean	SD	N
MVT	27.18	5.03	936	26.47	4.79	118
MCT	22.15	3.70	936	20.36	3.90	118
SAT	27.79	4.77	936	26.36	5.28	118
ANI	19.16	3.31	936	17.85	3.17	118
NSS	52.01	5.65	935	47.23	6.53	118

and Verbal Test, MCT stands for Mechanical Comprehension Test, SAT stands for Spatial Apperception Test, and ANI for Aviation and Nautical Interest. NSS stands for Navy Standard Score, a standardized score for overall performance in API. These five scores are given for two groups, (1) those who successfully graduated from Primary flight training, and (2) those who attrited from Primary flight training for any reason. These groups will be called the PASS group and the FAIL group respectively. The sample sizes for each group are also given in Table 1. The attrition rate for this sample is seen to be 118/1054 (11.2%). This is quite typical for the attrition rate in primary training which historically has hovered about the 10% mark. The mean score is higher for the PASS group than for the FAIL group on all five predictor variables. The standard deviations across the two groups seem to be roughly comparable.

Table 2 gives the sample variance-covariance matrix for the PASS group while Table 3 gives the same information on the FAIL group. The values along the diagonal of each matrix give the sample variances of the appropriate variables, i.e., the square of the standard deviations from Table 1.

Although the variance-covariance matrix is appropriate for the numerical computations described later, the sample correlations among the five predictor variables are more informative to our eyes. Tables 4 and 5 exhibit the pattern of correlations for the PASS group and the FAIL group, respectively. The correlations appear to be quite similar for the two groups.

Table 2: The sample variance-covariance matrix S_{Pass} for $N = 936$ subjects who passed primary flight training.

Test	MVT	MCT	SAT	ANI	NSS
MVT	25.335				
MCT	8.375	13.660			
SAT	2.156	4.204	22.790		
ANI	2.050	4.252	3.043	10.930	
NSS	9.760	6.616	3.822	3.825	31.947

Table 3: The sample variance-covariance matrix S_{Fail} for $N = 118$ subjects who failed primary flight training.

Test	MVT	MCT	SAT	ANI	NSS
MVT	22.987				
MCT	6.701	15.188			
SAT	-1.721	5.399	27.875		
ANI	1.851	3.756	2.347	10.079	
NSS	11.967	8.687	6.095	1.531	42.691

Table 4: The correlation matrix of the five predictor variables for the $N = 936$ subjects who passed primary flight training.

Test	MVT	MCT	SAT	ANI	NSS
MVT	1.000				
MCT	.450	1.000			
SAT	.090	.238	1.000		
ANI	.123	.348	.193	1.000	
NSS	.343	.317	.142	.205	1.000

Table 5: The correlation matrix of the five predictor variables for the $N = 118$ subjects who failed primary flight training.

Test	MVT	MCT	SAT	ANI	NSS
MVT	1.000				
MCT	.359	1.000			
SAT	-.068	.262	1.000		
ANI	.122	.304	.140	1.000	
NSS	.382	.341	.177	.074	1.000

Making the decision on a new candidate

How do you make the decision to predict a pass or a fail for any given candidate on the basis of the observed scores on the five predictor variables? The answer to this question is to use Bayesian Decision Theory (Berger [1], Coombs, Dawes and Tversky [2]). Within this approach, we try to find that decision which has the minimum expected loss with respect to the probability of passing or failing.

Expectation in statistical parlance is the same as the average, with the average defined for any generic discrete function $f(x)$ as,

$$E[f(x)] = \sum_{j=1}^n f(x_j) P(x_j). \quad (1)$$

The loss corresponds to the function $f(x)$. It depends on two arguments: (1) what actually happened, called the *state of nature*, θ ; and (2) the decision taken. The decision is sometimes called the *action*, and we will write a_k as the k th decision taken to avoid confusion with d_k . The uncertainty surrounds which state of nature, θ_j , will actually occur and so we represent this uncertainty with a discrete probability distribution, $P(\theta)$. In symbols, the expected loss is written as,

$$\text{expected loss of decision } a_k = \sum_{j=1}^n L(\theta_j, a_k) P(\theta_j). \quad (2)$$

The current situation has only a total of $n = 2$ possible true states of nature,

$\theta_1 \equiv$ actual pass

$\theta_2 \equiv$ actual fail

and two possible decisions that could be taken under each state,

$a_1 \equiv$ predict pass

$a_2 \equiv$ predict fail.

There are four possible losses corresponding to all possible combinations of the two decisions and the two states of nature. The losses for each decision under each state of nature are represented by the loss matrix shown in Table 6.

Table 6: The loss matrix for the decision problem of choosing candidates to enter primary flight training.

		Predict Pass	Predict Fail
		a_1	a_2
Actual Pass	θ_1	$L(\theta_1, a_1)$ 0	$L(\theta_1, a_2)$ C_1
Actual Fail	θ_2	$L(\theta_2, a_1)$ C_2	$L(\theta_2, a_2)$ 0

By convention, a loss of 0 is assigned to the two correct decisions, i.e., predicting that a candidate will pass when he or she actually does pass training, or predicting that the candidate will fail when she or he actually does fail training. A placeholder is inserted for the monetary costs incurred, C_1 and C_2 , for the two ways to make an

incorrect decision, that is, predicting that a candidate will pass when he or she actually fails training, or predicting fail when the candidate actually passes training.

Inserting these costs into Equation (2) results in

$$\text{expected loss } a_1 = L(\theta_1, a_1)P(\theta_1) + L(\theta_2, a_1)P(\theta_2) \quad (3)$$

$$\text{expected loss predicted pass} = [0 \times P(\text{Pass})] + [C_2 \times P(\text{Fail})] \quad (4)$$

$$= C_2 \times P(\text{Fail}) \quad (5)$$

$$\text{expected loss } a_2 = L(\theta_1, a_2)P(\theta_1) + L(\theta_2, a_2)P(\theta_2) \quad (6)$$

$$\text{expected loss predicted fail} = [C_1 \times P(\text{Pass})] + [0 \times P(\text{Fail})] \quad (7)$$

$$= C_1 \times P(\text{Pass}) \quad (8)$$

The decision rule itself is quite simple. If the expected loss of the predicted pass is less than or equal to the expected loss of the predicted fail, then predict pass, otherwise predict fail.

$$C_2 \times P(\text{Fail}) \leq C_1 \times P(\text{Pass}) \quad (9)$$

$$\frac{C_2}{C_1} \leq \frac{P(\text{Pass})}{P(\text{Fail})} \quad (10)$$

$$\text{If } \frac{P(\text{Pass})}{P(\text{Fail})} \geq \frac{C_2}{C_1} \quad (11)$$

Then Predict Pass

Else Predict Fail

Numerical examples of decision rule

This section presents some numerical examples of the decision rule as given in Equation (11). It also illustrates that the *probability* of passing needed to *predict* a pass will change with the costs associated with the incorrect decisions. For the first example, let the cost of training someone who later fails be greater than the cost of rejecting someone who would have passed training. Table 7 presents an example of such a situation.

Table 7: The loss matrix for the decision problem of choosing candidates to enter primary flight training when it is more expensive to train eventual failures than to reject some successful candidates.

		Predict Pass	Predict Fail
		a_1	a_2
Actual Pass	θ_1	$L(\theta_1, a_1)$ 0	$L(\theta_1, a_2)$ \$200,000
	θ_2	$L(\theta_2, a_1)$ \$1,000,000	$L(\theta_2, a_2)$ 0

In an ideal situation these costs would be determined through a detailed economic analysis conducted by experts in training and selection. Unfortunately, I am unfamiliar with any such analyses. The costs used in the

following examples were chosen by me as not unreasonable numbers to illustrate the mechanics of the decision rule. In any case, the purpose of this report is to provide the proper quantitative framework for making decisions about predicting a pass or a fail for an individual. The "correct" cost figures, when agreed upon, can be easily inserted into the loss matrix. In the PPS, the main utility of the loss matrix may be in allowing users to assess the effects of different decision thresholds in a "what-if" exercise.

Suppose that the probability of passing primary flight training based on the scores the candidate obtained on the five predictor variables is 80%. Therefore, the probability of failing is 20%.

$$\begin{aligned} \text{If } \frac{P(\text{Pass})}{P(\text{Fail})} &\geq \frac{C_2}{C_1} \\ \text{Then} &\quad \text{Predict Pass} \\ \text{Else} &\quad \text{Predict Fail} \\ \frac{P(\text{Pass})}{P(\text{Fail})} &= \frac{.80}{.20} \\ &= 4 \\ \frac{C_2}{C_1} &= 5 \\ 4 &\not\geq 5 \\ \text{Therefore,} &\quad \text{Predict Fail} \end{aligned}$$

In this case, a *probability* of passing equal to 80% is simply not high enough to commit to a decision to *predict* a pass because of the high cost of the wrong decision to let someone into training when they fail. If a candidate's probability of passing based on the test scores were raised to 90%, then

$$\frac{P(\text{Pass})}{P(\text{Fail})} \geq \frac{C_2}{C_1}$$

and the candidate would be admitted into training as a predicted pass.

As the ratio of losses for these incorrect decisions start to climb, then it becomes even harder for a candidate to get accepted into training. As the following numerical example illustrates, an increased ratio of C_2 to C_1 raises the threshold for acceptance into training even higher. Table 8 presents the scenario when training is very expensive, but the pool of qualified applicants wishing to be trained is large. In this altered situation,

$$\frac{P(\text{Pass})}{P(\text{Fail})} \geq 15$$

therefore, the probability of passing must be around 94% or greater for the decision rule to recommend acceptance into training. Going back to first principles, this means that the average loss for predicting a pass for any probability of passing less than 94% is greater than the average loss of predicting a fail.

Of course, the change in the decision criterion can work just as well in the opposite direction. If the losses for the two incorrect decisions were judged to be of equal value as in Table 9 below, then the ratio of costs would change to

$$\frac{C_2}{C_1} = 1$$

Table 8: The loss matrix for the decision problem of choosing candidates to enter primary flight training when training is very expensive but the pool of qualified applicants is large.

		Predict Pass	Predict Fail
		a_1	a_2
Actual Pass	θ_1	$L(\theta_1, a_1)$ 0	$L(\theta_1, a_2)$ \$100,000
Actual Fail	θ_2	$L(\theta_2, a_1)$ \$1,500,000	$L(\theta_2, a_2)$ 0

As soon as the ratio of the probability of passing to the probability of failing,

$$\frac{P(\text{Pass})}{P(\text{Fail})}$$

becomes greater than 1, a predicted pass results. A probability of passing equal to 50% or greater would be sufficient to predict a pass in this case .

Table 9: The loss matrix for the decision problem of choosing candidates to enter primary flight training when it is equally as bad to reject someone who could have passed as it is to admit someone who fails.

		Predict Pass	Predict Fail
		a_1	a_2
Actual Pass	θ_1	$L(\theta_1, a_1)$ 0	$L(\theta_1, a_2)$ \$500,000
Actual Fail	θ_2	$L(\theta_2, a_1)$ \$500,000	$L(\theta_2, a_2)$ 0

The derivation of the posterior odds

Except for the potential subjective or economic difficulty in assigning the costs, C_1 and C_2 , the calculation of this part of the decision rule is trivial, unlike that of the calculation of the ratio of probabilities. A great degree of non-trivial mathematics must be brought to bear in that case. Fortunately, the difficult part of the mathematics has already been worked out by Press [3] and Geisser [4]. In this section, we present the general outline for the formula of the ratio,

$$\frac{P(\text{Pass})}{P(\text{Fail})}$$

as needed for the decision rule. The ratio is commonly known as the "posterior odds." The formula used in the computer program to calculate the numerical results is also given at the end of this section.

Actually, $P(\text{Pass})$ or $P(\text{Fail})$ is too simplistic a notation for what is really a probability based on considerable information. That information consists of the scores obtained on the five predictor variables by the candidate we wish to classify. Thus, let D_{n+1} stand for these scores for the " $n + 1$ st" subject, i.e., the new candidate we wish to classify. Then D_n is the notation for the data in the database from the n students who also have scores for the predictor variables and for whom we know, as well, their success or failure in primary flight training.

Therefore, we now employ the more accurate notation of

$$P(\text{Pass}|D_{n+1}) \text{ and } P(\text{Fail}|D_{n+1})$$

as the posterior classification probability for passing and failing. By Bayes's Formula these two posterior probabilities are written as a function of the likelihood and the prior probabilities,

$$P(\text{Pass}|D_{n+1}) = \frac{P(D_{n+1}|\text{Pass}) P(\text{Pass})}{P(D_{n+1}|\text{Pass}) P(\text{Pass}) + P(D_{n+1}|\text{Fail}) P(\text{Fail})} \quad (12)$$

$$P(\text{Fail}|D_{n+1}) = \frac{P(D_{n+1}|\text{Fail}) P(\text{Fail})}{P(D_{n+1}|\text{Pass}) P(\text{Pass}) + P(D_{n+1}|\text{Fail}) P(\text{Fail})} \quad (13)$$

In forming the posterior odds, we can cancel out the denominator in the posterior probability of both terms. The posterior odds now can be written as a ratio of the likelihood times the prior probability for passing and failing,

$$\frac{P(\text{Pass}|D_{n+1})}{P(\text{Fail}|D_{n+1})} = \frac{P(D_{n+1}|\text{Pass}) P(\text{Pass})}{P(D_{n+1}|\text{Fail}) P(\text{Fail})} \quad (14)$$

For an introduction to the Bayesian approach as used in scientific inference, see Blower [5].

So far the development has been relatively straightforward, but now the difficult part of the mathematics intrudes. The likelihood terms in Equation (14) are called predictive densities in the Bayesian approach. The predictive density for either one of our two classification categories should be written as

$$P(D_{n+1}|\text{Pass}, D_n) \text{ or } P(D_{n+1}|\text{Fail}, D_n)$$

showing the dependence on the past data, D_n . The predictive density marginalizes over all the parameters used in the assumptions for how the scores are generated. In this case, the assumption is multivariate normality for the scores with two parameters, μ and Σ , where μ is the vector of population means for the scores and Σ is the population variance-covariance matrix of the scores. Marginalizing over these two parameters to find the predictive density yields,

$$P(D_{n+1}|\text{Pass}, D_n) = \iint P(D_{n+1}|\mu, \Sigma, \text{Pass}, D_n) P(\mu, \Sigma|\text{Pass}, D_n) d\mu_{Pass} d\Sigma_{Pass} \quad (15)$$

This is the product of the likelihood of the scores for the new candidate and the posterior probability of the parameters with this product taken over all values of the parameters.

Skipping over all the gory mathematical manipulations of Equation (15) (see Press [3], pp. 153–155 for the details), we show only the final result. The predictive probabilities for the scores of a new candidate (D_{n+1}) given the old data (D_n) and a particular category (Pass, Fail) are shown first in a general outline form and then in explicit detail.

$$P(D_{n+1}|\text{Pass}, D_n) = \text{term 1} \times \text{term 2} \quad (16)$$

$$P(D_{n+1}|\text{Fail}, D_n) = \text{term 3} \times \text{term 4} \quad (17)$$

$$\text{Posterior Odds} = \frac{P(\text{Pass}|D_{n+1})}{P(\text{Fail}|D_{n+1})} \quad (18)$$

$$= \frac{P(D_{n+1}|\text{Pass}, D_n) \times P(\text{Pass})}{P(D_{n+1}|\text{Fail}, D_n) \times P(\text{Fail})} \quad (19)$$

$$= \frac{\text{term 1} \times \text{term 2}}{\text{term 3} \times \text{term 4}} \times \frac{P(\text{Pass})}{P(\text{Fail})} \quad (20)$$

For constructing the computer program to calculate the posterior odds, the detailed formulas follow. Terms 1 and 3 are patterned alike as are terms 2 and 4.

$$\text{term 1} = |(N_{Fail} - 1) S_{Fail}|^{1/2} \Gamma\left(\frac{N_{Pass}}{2}\right) \Gamma\left(\frac{N_{Fail} - p}{2}\right) N_{Pass}(N_{Fail} + 1)^{p/2} \quad (21)$$

$$\text{term 2} = \left[1 + \left(\frac{N_{Fail}}{N_{Fail}^2 - 1} \right) (D_{n+1} - \bar{x}_{Fail})^T S_{Fail}^{-1} (D_{n+1} - \bar{x}_{Fail}) \right]^{N_{Fail}/2} \quad (22)$$

$$\text{term 3} = |(N_{Pass} - 1) S_{Pass}|^{1/2} \Gamma \left(\frac{N_{Fail}}{2} \right) \Gamma \left(\frac{N_{Pass} - p}{2} \right) N_{Fail} (N_{Pass} + 1)^{p/2} \quad (23)$$

$$\text{term 4} = \left[1 + \left(\frac{N_{Pass}}{N_{Pass}^2 - 1} \right) (D_{n+1} - \bar{x}_{Pass})^T S_{Pass}^{-1} (D_{n+1} - \bar{x}_{Pass}) \right]^{N_{Pass}/2} \quad (24)$$

Table 10 briefly describes the symbols used in Equations (21) through (24).

Table 10: Explanation for symbols in Equations (21) through (24).

Symbol		Explanation
$P(\text{Pass})$	$P(\text{Fail})$	Prior Probability of Passing or Failing
N_{Pass}	N_{Fail}	Sample sizes of Pass and Fail groups
p		Number of predictor variables
\bar{x}_{Pass}	\bar{x}_{Fail}	Sample means on p tests
S_{Pass}	S_{Fail}	Sample covariance matrices for p tests
D_{n+1}		Scores on predictor variables for new candidate
$(D_{n+1} - \bar{x}_{Pass})^T$	$(D_{n+1} - \bar{x}_{Fail})^T$	Transpose of difference between vector of new score and sample means
$ S_{Pass} $	$ S_{Fail} $	Determinant of sample covariance matrices
S_{Pass}^{-1}	S_{Fail}^{-1}	Inverse of sample covariance matrices
$\Gamma(N)$		Gamma function with N as argument

Statistical Decision Theory

Before beginning the numerical exercises, we would like to express the equations just derived in a form compatible with Bayesian statistical decision theory. The decision algorithm from statistical decision theory can be stated in the following simple terminology. If the likelihood ratio is greater than some response threshold, then commit to a particular decision. In symbols,

If $\mathcal{L}(x) \geq \beta$ then predict appropriate outcome

Equation (11), as the ratio of the posterior odds, should now be correctly written as

$$\text{If } \frac{P(\text{Pass}|D_{n+1})}{P(\text{Fail}|D_{n+1})} \geq \frac{C_2}{C_1} \text{ then Predict Pass otherwise Predict Fail} \quad (25)$$

Substituting Equation (19) for the posterior odds yields,

$$\frac{P(D_{n+1}|\text{Pass}, D_n) \times P(\text{Pass})}{P(D_{n+1}|\text{Fail}, D_n) \times P(\text{Fail})} \geq \frac{C_2}{C_1} \quad (26)$$

$$\frac{P(D_{n+1}|\text{Pass}, D_n)}{P(D_{n+1}|\text{Fail}, D_n)} \geq \frac{C_2}{C_1} \times \frac{P(\text{Fail})}{P(\text{Pass})} \quad (27)$$

The left hand side of Equation (27) is in the form of a likelihood ratio since it is the ratio of the probability of the data from a candidate given the old data and the Pass group over the probability of the data from a candidate given the old data and the Fail group. That is,

$$\mathcal{L}(x) \equiv \mathcal{L}(\text{data from candidate})$$

The right hand side of Equation (27) is a function of the costs of making correct and incorrect decisions and the prior odds of failing over passing. Together they make up β , the response threshold. Equation (27) therefore represents a decision algorithm in the form of

$$\text{If } \mathcal{L}(x) \geq \beta \text{ then predict Pass}$$

Numerical computations

Numerical examples of the Bayesian decision rule as derived in the previous sections are presented here. Table 11 shows the scores obtained on the five predictor variables for three new candidates for whom we would like a classification as a predicted pass or a predicted fail. The criterion for such a prediction will be to minimize the monetary loss experienced with an incorrect decision. For this first set of examples, we arbitrarily set the ratio C_2/C_1 equal to 5, merely as a reasonable supposition about costs. For example, an incorrect rejection of a candidate might cost \$100,000, and permitting an eventual attrition into primary flight training might cost \$500,000. Again, the five predictor variables are the scores on four subtests of the ASTB, the MVT, MCT, SAT, and ANI, and the final composite grade from API, NSS. According to historical records, it is known that the probability of passing primary flight training given the current selection standards, but using no additional information as we are doing here, is about 90%. Therefore, we will set the prior probability of passing at 90% and the prior probability of failing at 10%. The posterior classification probabilities are calculated by using Equations (21) through (24). The sample means are taken from Table 1 and the sample covariances from Tables 2 and 3. The sample size for the Pass group is $N_{Pass} = 936$ and the sample size for the Fail group is $N_{Fail} = 118$. The number of predictor variables, p , is equal to 5. The D_{n+1} are the scores listed in the five columns of Table 11 after the *Candidate* column.

Table 11: Examples of Bayesian decision rule for various scores obtained by three candidates. The loss ratio for this example is set at $C_2/C_1 = 5$.

Candidate	MVT	MCT	SAT	ANI	NSS	$P(\text{Pass} D_{n+1})$	Post. Odds	Decision
1	27	22	27	19	52	.9413	16.037	Predicted Pass
2	26	20	26	17	47	.8744	6.963	Predicted Pass
3	24	18	23	15	40	.6608	1.948	Predicted Fail

The first candidate scored close to the sample mean of the 936 students who passed primary flight training. It is not surprising, therefore, that the posterior classification probability is quite high (at about 94%) that this

candidate will pass training. Even so, a high probability of passing training is not, in and of itself, enough to make a decision to predict a pass. For this candidate though, the posterior odds of 16.037 are greater than 5 so we can safely predict pass given our cost structure. The second candidate had lower scores on all five predictor variables and the associated posterior classification probability is correspondingly lower at around 87%. However, the posterior odds for this candidate are still greater than 5, so we again predict a pass. Had the cost structure been different, say $C_2/C_1 = 10$, in order to emphasize the importance of admitting into training only those with a very high probability of success, then a different decision would be warranted for this candidate. The third candidate achieved even lower scores to the point where the posterior classification probability of passing is lowered to 66%. Now the posterior odds drop below the threshold of 5, and the decision is made to predict that this candidate will fail.

Effect of varying parameters

It is easy to examine the effect on the probability of passing primary flight training by changing any of the parameters of the quantitative model. An example is provided of how this can be done by making the following changes:

- [1] decreasing the size of the number of students who passed and failed training
- [2] changing the prior probability of passing and failing
- [3] making the tests more discriminatory by widening the gap in the sample means
- [4] increasing the correlation among the predictor variables.

These four changes are operationalized within the quantitative model by setting the values of the following parameters:

- [1] $N_{Pass} = 400$ and $N_{Fail} = 100$
- [2] $P(Pass) = .80$ and $P(Fail) = .20$
- [3] $\bar{x}_{Pass} = (28, 23, 28, 20, 55)$ and $\bar{x}_{Fail} = (25, 19, 24, 16, 45)$
- [4] increasing the off-diagonal elements of S_{Pass} and S_{Fail} .

The effect of changing the cost structure on incorrect decisions has already been discussed, therefore this facet of the decision rule will be kept the same as in the last numerical example where $C_2/C_1 = 5$. The numerical results from the changes just discussed are given in Table 12. Like Table 11, Table 12 gives the scores of three

Table 12: The effect on the probability of passing and the decision to predict pass or fail from changing parameters in the quantitative model.

Candidate	MVT	MCT	SAT	ANI	NSS	$P(Pass D_{n+1})$	Post. Odds	Decision
1	28	23	28	20	55	.9747	38.487	Predicted Pass
2	26	20	25	17	47	.5109	1.045	Predicted Fail
3	28	23	28	20	45	.6088	1.556	Predicted Fail

candidates on the five predictor variables. As in the last example, the first candidate scores at the mean of the pass group, but since the sample means are further apart than in the first example, these scores are now more compelling evidence that this candidate will pass. The posterior odds are 38.487:1 as opposed to 16.037:1 for the first example. The second candidate scores better on all tests in that he is above the mean of the Fail group, whereas his counterpart of the first example was only at the mean of the Fail group. This candidate is a predicted fail whereas the candidate in the first example was a predicted pass. Again, this is because the means are more

widely separated, with the result that this second test battery is more discriminating. Another contributing factor to the predicted fail is the lower prior probability of a pass in this example.

A curious phenomenon is revealed by the third candidate. For the sake of discussion, this candidate happens to be female. She scores exactly the same as the first candidate on the four ASTB tests, but is at the mean of the fail group for the API score. The first candidate was a confident predicted pass, but this candidate is an equally resounding predicted fail. What happened here? Is the result of just one test score so detrimental to this candidate's prospects?

The answer lies in the nature of the changed covariance matrices. The off-diagonal elements of the covariance matrices for both the pass and fail groups were made larger to reflect a situation where the correlation among all five tests was extremely high. For example, the true correlation between MVT and MCT was set at .90, and the true correlation between MVT and NSS was set at .70. The pattern of intercorrelations for the first four scores with the fifth score demands a score close to the mean of the pass group to fit the profile of a pass candidate. The discordant score on the last predictor variable is enough, given the extremely high intercorrelations, to make her much less like the typical pass candidate. The posterior odds are correspondingly lowered and given the costs of the various decisions make her a predicted fail. The fact that the prior probability of a pass was lowered did not help this candidate either.

Another way of looking at this situation of the third candidate is in terms of the number of pieces of independent information. In the ideal case, there would be no correlation among the five predictor variables and the off diagonal elements of the sample variance-covariance matrices would be small. Each of the five test scores would then be providing five independent sources of information about the correct category the candidate belongs to. In the situation, however, of large off-diagonal elements in the sample variance-covariance matrices, the test scores are highly correlated. There are not really five pieces of independent information. For the third candidate we could say roughly that there are only two pieces of independent information, one indicating she belongs to the pass group and the other indicating that she belongs to the fail group. So, on this viewpoint, it is not surprising then that her probability of passing is driven down towards 50%.

Summary

The high-level policy decisions that depend on access to selection and training data could be improved by a more coordinated effort at upgrading data bases and integrating statistical prediction models. NAMRL believes that the Pilot Prediction System is a research effort that will move us closer to that goal.

This report outlines some initial thoughts on the architecture of the PPS and how it might help its targeted audience. More specifically, we have presented in some detail a statistical model that predicts whether a student will pass or fail in primary flight training as a function of four selection test scores and overall achievement in API. It not only deals with this kind of quantitative information in an optimal fashion through the Bayesian assessment of the probability of passing, but also takes into account the judgmental or economic cost factors involved in making correct and incorrect decisions.

A computer program has been written to implement the Bayesian posterior classification probability as well as the decision theory aspects. Some numerical examples were presented to show how the program works to predict success for candidates who have not yet entered primary flight training, but for whom we have scores on five predictor variables. Such a program can serve as one core module in a user friendly, spreadsheet-like implementation of the PPS. The PPS implementation philosophy is to shield its users from technical issues such as using statistical prediction models to extrapolate from success profiles in a data base. This feature permits them to concentrate more fully on the larger manpower and training issues that beset naval aviation.

References

1. Berger, J. O. *Statistical Decision Theory and Bayesian Analysis*. Springer-Verlag, Berlin, Germany, 1985.
2. Coombs, C. H., Dawes, R. M., and Tversky, A. *Mathematical Psychology: An Elementary Introduction*. Prentice-Hall, Englewood Cliffs, NJ, 1970.
3. Press, S. J. *Bayesian statistics: principles, models, and applications*. John Wiley & Sons, New York, NY, 1989.
4. Geisser, S. *An Introduction to Predictive Inference*. Chapman and Hall, New York, NY, 1993.
5. Blower, D. J. *An Introduction to Scientific Inference: Volume I The Bayesian Fundamentals*. Published privately by the author, 1998. Individual chapters can be downloaded from web site <http://www.namrl.navy.mil>.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503.

1. AGENCY USE ONLY (Leave blank)		2. REPORT DATE 30 Nov 1998	3. REPORT TYPE AND DATES COVERED	
4. TITLE AND SUBTITLE Probability of Success in Primary Flight Training as a Function of ASTB Scores and API Grades: An Example of the Statistical Inferencing Component of the Pilot Prediction System.			5. FUNDING NUMBERS 62233N MM33P30.001-7801	
6. AUTHOR(S) David J. Blower				
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Naval Aerospace Medical Research Laboratory 51 Hovey Road Pensacola FL 32508-1046			8. PERFORMING ORGANIZATION REPORT NUMBER NAMRL-1404	
9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES) Naval Medical Research and Development Command National Naval Medical Center Bldg. 1, Tower 12 8901 Wisconsin Avenue Bethesda, MD 20889-5606			10. SPONSORING / MONITORING AGENCY REPORT NUMBER	
11. SUPPLEMENTARY NOTES				
12a. DISTRIBUTION / AVAILABILITY STATEMENT Approved for public release; distribution unlimited.			12b. DISTRIBUTION CODE	
13. ABSTRACT (Maximum 200 words) The Pilot Prediction System (PPS) is a research effort designed to provide Navy managers and other decision makers with improved access to selection and training data. Many disparate data bases, each containing partial and sometimes overlapping information on selection data and training performance, currently exist. There has been no attempt to coordinate the bits and pieces gathered into these local databases into a coherent whole. Such data needs to be merged and the anomalies excised so that a more global picture of selection data and training performance can emerge. In addition, the targeted users of the PPS should be shielded from the low-level technicalities of the data base because such technical details are of no concern to them. For the same reason, the statistical manipulations that provide extrapolations from the data base to new cases can be hidden from view. This report documents the first efforts at constructing the statistical modeling component of the PPS as derived from Bayesian statistical decision theory. It enables the potential user of the PPS to predict success in primary flight training for flight students based on their scores on selection tests and ground school performance.				
14. SUBJECT TERMS Selection and training; Training prediction; Pilot prediction system; Bayesian decision theory			15. NUMBER OF PAGES 19	
			16. PRICE CODE	
17. SECURITY CLASSIFICATION OF REPORT UNCLASSIFIED	18. SECURITY CLASSIFICATION OF THIS PAGE UNCLASSIFIED	19. SECURITY CLASSIFICATION OF ABSTRACT UNCLASSIFIED	20. LIMITATION OF ABSTRACT SAR	